

Paper #6

S. Baum

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/675,208

DATE: 11/30/2001

TIME: 11:43:22

Input Set : N:\Crf3\RULE60\09675208.txt

Output Set: N:\CRF3\11212001\I675208.raw

## SEQUENCE LISTING

3 (1) GENERAL INFORMATION:

4 (i) APPLICANT: OSUMI Chieko

5 NOZAKI Jinshi

6 KIDA Takao

7 (ii) TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR

8 PRODUCING RAFFINOSE, AND TRANSGENIC PLANT

9 (iii) NUMBER OF SEQUENCES: 22

10 (iv) CORRESPONDENCE ADDRESS:

11 (A) ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.

12 (B) STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR

13 (C) CITY: ARLINGTON

14 (D) STATE: VIRGINIA

15 (E) COUNTRY: USA

16 (F) ZIP: 22202

17 (v) COMPUTER READABLE FORM:

18 (A) MEDIUM TYPE: Floppy disk

19 (B) COMPUTER: IBM PC compatible

20 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

21 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

22 (vi) CURRENT APPLICATION DATA:

C--> 23 (A) APPLICATION NUMBER: US/09/675,208

C--> 24 (B) FILING DATE: 29-Sep-2000

25 (vii) PRIOR APPLICATION DATA:

26 (A) APPLICATION NUMBER: 08/846,234

27 (B) FILING DATE:

C--> 28 (viii) ATTORNEY/AGENT INFORMATION:

29 (A) NAME: NORMAN F. OBLON

30 (B) REGISTRATION NUMBER: 24,618

31 (ix) TELECOMMUNICATION INFORMATION:

32 (A) TELEPHONE: (703)-413-3000

33 (B) TELEFAX: (703)-413-2220

34 (2) INFORMATION FOR SEQ ID NO: 1:

35 (i) SEQUENCE CHARACTERISTICS:

36 (A) LENGTH: 30 amino acids

37 (B) TYPE: amino acid

38 (D) TOPOLOGY: linear

39 (ii) MOLECULE TYPE: peptide

40 (v) FRAGMENT TYPE: internal

41 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

42 Phe Gly Trp Cys Thr Trp Asp Ala Phe Tyr Leu Thr Val His Pro Gln

43 1 5 10 15

44 Gly Val Ile Glu Gly Val Arg His Leu Val Asp Gly Gly Cys

45 20 25 30

46 (2) INFORMATION FOR SEQ ID NO: 2:

47 (i) SEQUENCE CHARACTERISTICS:

48 (A) LENGTH: 19 amino acids

ENTERED

## RAW SEQUENCE LISTING

DATE: 11/30/2001

PATENT APPLICATION: US/09/675,208

TIME: 11:43:22

Input Set : N:\Crf3\RULE60\09675208.txt

Output Set: N:\CRF3\11212001\I675208.raw

```

56          (B) TYPE: amino acid
57          (D) TOPOLOGY: linear
58      (ii) MOLECULE TYPE: peptide
59      (v) FRAGMENT TYPE: internal
60      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
61 Pro Val Ser Val Gly Cys Phe Val Gly Phe Asp Ala Ser Glu Pro Asp
62  1          5          10          15
63 Ser Arg His
64 (2) INFORMATION FOR SEQ ID NO: 3:
65      (i) SEQUENCE CHARACTERISTICS:
66          (A) LENGTH: 14 amino acids
67          (B) TYPE: amino acid
68          (D) TOPOLOGY: linear
69      (ii) MOLECULE TYPE: peptide
70      (v) FRAGMENT TYPE: internal
71      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
72 Tyr Asp Gln Asp Gln Met Val Val Val Gln Val Pro Trp Pro
73  1          5          10
74 (2) INFORMATION FOR SEQ ID NO: 4:
75      (i) SEQUENCE CHARACTERISTICS:
76          (A) LENGTH: 2517 base pairs
77          (B) TYPE: nucleic acid
78          (C) STRANDEDNESS: double
79          (D) TOPOLOGY: linear
80      (ii) MOLECULE TYPE: cDNA to mRNA
81      (vi) ORIGINAL SOURCE:
82          (A) ORGANISM: cucumber (Cucumis sativas)
83      (ix) FEATURE:
84          (A) NAME/KEY: CDS
85          (B) LOCATION: 56..2407
86      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
87 AAAAAACAAC CCTTCTTTTA GTTTTTTGGG TTTGTTTCTT CTTTCTTCT CACAA ATG      58
88                                         Met
89                                         1
90 GCT CCT AGT TTT AAA AAT GGT GGC TCC AAC GTA GTT TCA TTT GAT GGC      106
91 Ala Pro Ser Phe Lys Asn Gly Gly Ser Asn Val Val Ser Phe Asp Gly
92          5          10          15
93 TTA AAT GAC ATG TCG TCA CCG TTT GCA ATC GAC GGA TCG GAT TTC ACT      154
94 Leu Asn Asp Met Ser Ser Pro Phe Ala Ile Asp Gly Ser Asp Phe Thr
95          20          25          30
96 GTG AAC GGT CAT TCG TTT CTG TCC GAT GTT CCT GAG AAC ATT GTT GCT      202
97 Val Asn Gly His Ser Phe Leu Ser Asp Val Pro Glu Asn Ile Val Ala
98          35          40          45
99 TCT CCT TCT CCG TAC ACT TCG ATA GAC AAG TCC CCG GTT TCG GTT GGT      250
100 Ser Pro Ser Pro Tyr Thr Ser Ile Asp Lys Ser Pro Val Ser Val Gly
101          50          55          60          65
102 TGC TTT GTT GGA TTC GAC GCG TCG GAA CCT GAT AGC CGA CAT GTT GTT      298
103 Cys Phe Val Gly Phe Asp Ala Ser Glu Pro Asp Ser Arg His Val Val
104          70          75          80

```

## RAW SEQUENCE LISTING

DATE: 11/30/2001

PATENT APPLICATION: US/09/675,208

TIME: 11:43:22

Input Set : N:\Crf3\RULE60\09675208.txt

Output Set: N:\CRF3\11212001\I675208.raw

107	TCG	ATT	GGG	AAG	CTG	AAG	GAT	ATT	CGG	TTT	ATG	AGT	ATT	TTC	AGG	TTT	346
108	Ser	Ile	Gly	Lys	Leu	Lys	Asp	Ile	Arg	Phe	Met	Ser	Ile	Phe	Arg	Phe	
109			85						90					95			
110	AAG	GTT	TGG	TGG	ACT	ACA	CAC	TGG	GTT	GGT	CGA	AAT	GGT	GGG	GAT	CTT	394
111	Lys	Val	Trp	Trp	Thr	Thr	His	Trp	Val	Gly	Arg	Asn	Gly	Gly	Asp	Leu	
112			100					105					110				
113	GAA	TCG	GAG	ACT	CAG	ATT	GTG	ATC	CTT	GAG	AAG	TCA	GAT	TCT	GGT	CGA	442
114	Glu	Ser	Glu	Thr	Gln	Ile	Val	Ile	Leu	Glu	Lys	Ser	Asp	Ser	Gly	Arg	
115		115					120			125							
116	CCG	TAT	GTT	TTC	CTT	CTT	CCG	ATC	GTT	GAG	GGA	CCG	TTC	CGA	ACC	TCG	490
117	Pro	Tyr	Val	Phe	Leu	Leu	Pro	Ile	Val	Glu	Gly	Pro	Phe	Arg	Thr	Ser	
118	130					135				140					145		
119	ATT	CAG	CCT	GGG	GAT	GAT	GAC	TTT	GTC	GAT	GTT	TGT	GTC	GAG	AGT	GGT	538
120	Ile	Gln	Pro	Gly	Asp	Asp	Asp	Phe	Val	Asp	Val	Cys	Val	Glu	Ser	Gly	
121			150						155				160				
122	TCG	TCG	AAA	GTT	GTT	GAT	GCA	TCG	TTC	CGA	AGT	ATG	TTG	TAT	CTT	CAT	586
123	Ser	Ser	Lys	Val	Val	Asp	Ala	Ser	Phe	Arg	Ser	Met	Leu	Tyr	Leu	His	
124			165					170					175				
125	GCT	GGT	GAT	GAT	CCG	TTT	GCA	CTT	GTT	AAA	GAG	GCG	ATG	AAG	ATC	GTG	634
126	Ala	Gly	Asp	Asp	Pro	Phe	Ala	Leu	Val	Lys	Glu	Ala	Met	Lys	Ile	Val	
127		180					185				190						
128	AGG	ACC	CAT	CTT	GGA	ACT	TTT	CGC	TTG	TTG	GAG	GAG	AAG	ACT	CCA	CCA	682
129	Arg	Thr	His	Leu	Gly	Thr	Phe	Arg	Leu	Leu	Glu	Glu	Lys	Thr	Pro	Pro	
130		195				200				205							
131	GGT	ATC	GTG	GAC	AAA	TTC	GGT	TGG	TGC	ACG	TGG	GAC	GCG	TTT	TAC	CTA	730
132	Gly	Ile	Val	Asp	Lys	Phe	Gly	Trp	Cys	Thr	Trp	Asp	Ala	Phe	Tyr	Leu	
133	210				215				220				225				
134	ACG	GTT	CAT	CCA	CAG	GGC	GTA	ATA	GAA	GGC	GTG	AGG	CAT	CTC	GTC	GAC	778
135	Thr	Val	His	Pro	Gln	Gly	Val	Ile	Glu	Gly	Val	Arg	His	Leu	Val	Asp	
136			230					235					240				
137	GGC	GGT	TGT	CCT	CCC	GGT	TTA	GTC	CTA	ATC	GAC	GAT	GGT	TGG	CAA	TCC	826
138	Gly	Gly	Cys	Pro	Pro	Gly	Leu	Val	Leu	Ile	Asp	Asp	Gly	Trp	Gln	Ser	
139			245				250						255				
140	ATC	GGA	CAC	GAT	TCG	GAT	CCC	ATC	ACC	AAA	GAA	GGA	ATG	AAC	CAA	ACC	874
141	Ile	Gly	His	Asp	Ser	Asp	Pro	Ile	Thr	Lys	Glu	Gly	Met	Asn	Gln	Thr	
142		260					265						270				
143	GTC	GCC	GGC	GAG	CAA	ATG	CCC	TGC	CGT	CTT	TTG	AAA	TTC	CAA	GAG	AAT	922
144	Val	Ala	Gly	Glu	Gln	Met	Pro	Cys	Arg	Leu	Leu	Lys	Phe	Gln	Glu	Asn	
145		275				280				285							
146	TAC	AAA	TTC	CGT	GAC	TAC	GTC	AAT	CCC	AAG	GCC	ACC	GGC	CCC	CGA	GCC	970
147	Tyr	Lys	Phe	Arg	Asp	Tyr	Val	Asn	Pro	Lys	Ala	Thr	Gly	Pro	Arg	Ala	
148	290				295				300				305				
149	GGC	CAG	AAG	GGG	ATG	AAG	GCG	TTT	ATA	GAT	GAA	CTC	AAA	GGA	GAG	TTT	1018
150	Gly	Gln	Lys	Gly	Met	Lys	Ala	Phe	Ile	Asp	Glu	Leu	Lys	Gly	Glu	Phe	
151			310					315					320				
152	AAG	ACT	GTG	GAG	CAT	GTT	TAT	GTT	TGG	CAT	GCT	TTG	TGT	GGA	TAT	TGG	1066
153	Lys	Thr	Val	Glu	His	Val	Tyr	Val	Trp	His	Ala	Leu	Cys	Gly	Tyr	Trp	
154			325				330						335				
155	GGT	GGC	CTT	CGC	CCG	CAG	GTG	CCT	GGC	TTG	CCT	GAG	GCA	CGT	GTG	ATT	1114

## RAW SEQUENCE LISTING

DATE: 11/30/2001

PATENT APPLICATION: US/09/675,208

TIME: 11:43:22

Input Set : N:\Crf3\RULE60\09675208.txt

Output Set: N:\CRF3\11212001\I675208.raw

156	Gly	Gly	Leu	Arg	Pro	Gln	Val	Pro	Gly	Leu	Pro	Glu	Ala	Arg	Val	Ile	
157			340					345					350				
158	CAG	CCA	GTG	CTT	TCA	CCA	GGG	CTG	CAG	ATG	ACG	ATG	GAG	GAT	TTG	GCG	1162
159	Gln	Pro	Val	Leu	Ser	Pro	Gly	Leu	Gln	Met	Thr	Met	Glu	Asp	Leu	Ala	
160			355				360					365					
161	GTG	GAT	AAG	ATT	GTT	CTT	CAT	AAG	GTC	GGG	CTG	GTC	CCG	CCG	GAG	AAG	1210
162	Val	Asp	Lys	Ile	Val	Leu	His	Lys	Val	Gly	Leu	Val	Pro	Pro	Glu	Lys	
163	370					375				380					385		
164	GCT	GAG	GAG	ATG	TAC	GAA	GGA	CTT	CAT	GCT	CAT	TTG	GAA	AAA	GTT	GGG	1258
165	Ala	Glu	Glu	Met	Tyr	Glu	Gly	Leu	His	Ala	His	Leu	Glu	Lys	Val	Gly	
166				390				395						400			
167	ATC	GAC	GGT	GTT	AAG	ATT	GAC	GTT	ATC	CAC	CTA	TTG	GAG	ATG	TTG	TGT	1306
168	Ile	Asp	Gly	Val	Lys	Ile	Asp	Val	Ile	His	Leu	Leu	Glu	Met	Leu	Cys	
169			405					410					415				
170	GAA	GAC	TAT	GGA	GGG	AGA	GTG	GAT	TTG	GCA	AAG	GCA	TAT	TAC	AAA	GCA	1354
171	Glu	Asp	Tyr	Gly	Gly	Arg	Val	Asp	Leu	Ala	Lys	Ala	Tyr	Tyr	Lys	Ala	
172			420					425					430				
173	ATG	ACC	AAA	TCA	ATA	AAT	AAA	CAT	TTT	AAA	GGA	AAT	GGA	GTC	ATT	GCA	1402
174	Met	Thr	Lys	Ser	Ile	Asn	Lys	His	Phe	Lys	Gly	Asn	Gly	Val	Ile	Ala	
175			435				440					445					
176	AGT	ATG	GAA	CAT	TGT	AAC	GAC	TTC	ATG	TTC	CTT	GGC	ACG	GAA	GCT	ATC	1450
177	Ser	Met	Glu	His	Cys	Asn	Asp	Phe	Met	Phe	Leu	Gly	Thr	Glu	Ala	Ile	
178	450					455				460					465		
179	TCT	CTT	GGT	CGT	GTT	GGT	GAT	GAC	TTT	TGG	TGC	ACG	GAC	CCC	TCT	GGT	1498
180	Ser	Leu	Gly	Arg	Val	Gly	Asp	Asp	Phe	Trp	Cys	Thr	Asp	Pro	Ser	Gly	
181				470				475							480		
182	GAT	CCA	AAC	GGT	ACG	TTT	TGG	CTC	CAA	GGA	TGT	CAC	ATG	GTT	CAT	TGT	1546
183	Asp	Pro	Asn	Gly	Thr	Phe	Trp	Leu	Gln	Gly	Cys	His	Met	Val	His	Cys	
184				485				490					495				
185	GCC	AAC	GAC	AGC	TTG	TGG	ATG	GGG	AAC	TTC	ATC	CAC	CCT	GAC	TGG	GAT	1594
186	Ala	Asn	Asp	Ser	Leu	Trp	Met	Gly	Asn	Phe	Ile	His	Pro	Asp	Trp	Asp	
187			500					505					510				
188	ATG	TTC	CAA	TCC	ACC	CAC	CCT	TGT	GCC	GCC	TTC	CAT	GCT	GCC	TCT	CGA	1642
189	Met	Phe	Gln	Ser	Thr	His	Pro	Cys	Ala	Ala	Phe	His	Ala	Ala	Ser	Arg	
190			515				520					525					
191	GCC	ATC	TCT	GGT	GGC	CCG	ATC	TAT	GTT	AGT	GAT	TCT	GTG	GGA	AAG	CAT	1690
192	Ala	Ile	Ser	Gly	Gly	Pro	Ile	Tyr	Val	Ser	Asp	Ser	Val	Gly	Lys	His	
193	530					535				540					545		
194	AAC	TTT	GAT	CTT	CTG	AAA	AAA	CTA	GTG	CTT	CCT	GAT	GGA	TCG	ATC	CTT	1738
195	Asn	Phe	Asp	Leu	Leu	Lys	Lys	Leu	Val	Leu	Pro	Asp	Gly	Ser	Ile	Leu	
196				550				555					560				
197	CGA	AGT	GAG	TAC	TAT	GCA	CTC	CCG	ACT	CGC	GAT	TGT	TTG	TTT	GAA	GAC	1786
198	Arg	Ser	Glu	Tyr	Tyr	Ala	Leu	Pro	Thr	Arg	Asp	Cys	Leu	Phe	Glu	Asp	
199				565				570					575				
200	CCT	TTG	CAT	AAT	GGA	GAA	ACT	ATG	CTT	AAG	ATT	TGG	AAT	CTC	AAC	AAG	1834
201	Pro	Leu	His	Asn	Gly	Glu	Thr	Met	Leu	Lys	Ile	Trp	Asn	Leu	Asn	Lys	
202			580					585					590				
203	TTC	ACT	GGA	GTG	ATT	GGT	GCA	TTC	AAC	TGC	CAA	GGA	GGA	GGA	TGG	TGT	1882
204	Phe	Thr	Gly	Val	Ile	Gly	Ala	Phe	Asn	Cys	Gln	Gly	Gly	Gly	Trp	Cys	

## RAW SEQUENCE LISTING

DATE: 11/30/2001

PATENT APPLICATION: US/09/675,208

TIME: 11:43:22

Input Set : N:\Crf3\RULE60\09675208.txt

Output Set: N:\CRF3\11212001\I675208.raw

```

205      595      600      605
206 CGT GAG ACA CGC CGC AAC CAA TGC TTT TCA CAA TAC TCA AAA CGA GTG      1930
207 Arg Glu Thr Arg Arg Asn Gln Cys Phe Ser Gln Tyr Ser Lys Arg Val
208 610      615      620      625
209 ACA TCC AAA ACT AAC CCA AAA GAC ATA GAA TGG CAC AGT GGA GAA AAC      1978
210 Thr Ser Lys Thr Asn Pro Lys Asp Ile Glu Trp His Ser Gly Glu Asn
211      630      635      640
212 CCT ATC TCT ATT GAA GGC GTT AAA ACC TTT GCG CTT TAC CTC TAT CAA      2026
213 Pro Ile Ser Ile Glu Gly Val Lys Thr Phe Ala Leu Tyr Leu Tyr Gln
214      645      650      655
215 GCC AAA AAA CTT ATC CTC TCC AAG CCC TCT CAA GAT CTT GAC ATA GCT      2074
216 Ala Lys Lys Leu Ile Leu Ser Lys Pro Ser Gln Asp Leu Asp Ile Ala
217      660      665      670
218 CTT GAC CCA TTC GAA TTC GAG CTC ATC ACT GTT TCA CCA GTG ACC AAA      2122
219 Leu Asp Pro Phe Glu Phe Glu Leu Ile Thr Val Ser Pro Val Thr Lys
220      675      680      685
221 CTC ATC CAA ACT TCT CTA CAC TTT GCC CCA ATT GGG CTG GTG AAC ATG      2170
222 Leu Ile Gln Thr Ser Leu His Phe Ala Pro Ile Gly Leu Val Asn Met
223 690      695      700      705
224 CTT AAC ACT AGT GGA GCC ATC CAA TCT GTG GAC TAT GAC GAT GAC CTA      2218
225 Leu Asn Thr Ser Gly Ala Ile Gln Ser Val Asp Tyr Asp Asp Asp Leu
226      710      715      720
227 AGC TCA GTC GAG ATT GGT GTC AAA GGG TGT GGT GAG ATG CGA GTA TTT      2266
228 Ser Ser Val Glu Ile Gly Val Lys Gly Cys Gly Glu Met Arg Val Phe
229      725      730      735
230 GCA TCG AAA AAA CCA AGG GCT TGT CGT ATT GAT GGG GAG GAT GTT GGG      2314
231 Ala Ser Lys Lys Pro Arg Ala Cys Arg Ile Asp Gly Glu Asp Val Gly
232      740      745      750
233 TTC AAG TAT GAT CAG GAC CAA ATG GTG GTG GTT CAA GTG CCA TGG CCA      2362
234 Phe Lys Tyr Asp Gln Asp Gln Met Val Val Val Gln Val Pro Trp Pro
235      755      760      765
236 ATT GAT TCT TCA TCG GGT GGC ATT TCG GTT ATC GAG TAC TTG TTT      2407
237 Ile Asp Ser Ser Ser Gly Gly Ile Ser Val Ile Glu Tyr Leu Phe
238 770      775      780
239 TAATTTTTTAT TTATGTAAGC TCAATGATTG TTGTTGTTGT CGCTGTTGTT GCTATCAATG      2467
240 TATTTCTCTC CAAAAGAAAA TTATGTGTAA TTTGGAGAGT AATTAAGTGA      2517
242 (2) INFORMATION FOR SEQ ID NO: 5:
243 (i) SEQUENCE CHARACTERISTICS:
244 (A) LENGTH: 784 amino acids
245 (B) TYPE: amino acid
246 (D) TOPOLOGY: linear
247 (ii) MOLECULE TYPE: protein
248 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
249 Met Ala Pro Ser Phe Lys Asn Gly Gly Ser Asn Val Val Ser Phe Asp
250 1 5 10 15
251 Gly Leu Asn Asp Met Ser Ser Pro Phe Ala Ile Asp Gly Ser Asp Phe
252 20 25 30
253 Thr Val Asn Gly His Ser Phe Leu Ser Asp Val Pro Glu Asn Ile Val
254 35 40 45

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/675,208

DATE: 11/30/2001

TIME: 11:43:23

Input Set : N:\Crf3\RULE60\09675208.txt

Output Set: N:\CRF3\11212001\I675208.raw

L:24 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:25 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]

L:32 M:220 C: Keyword misspelled or invalid format, [(viii) ATTORNEY/AGENT INFORMATION:]